



Defence Research and  
Development Canada

Recherche et développement  
pour la défense Canada

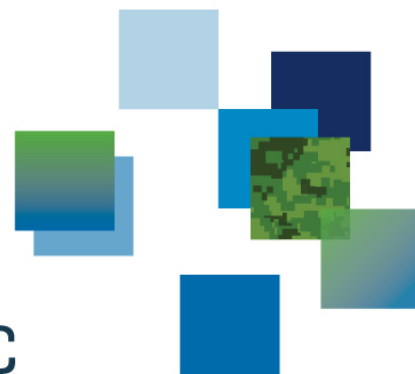
# Differential Impacts of Demographics in Biometric Systems

Dr. John W. M. Campbell

NIST IFPC

November 27, 2018

Contact: [john@bionbiometrics.com](mailto:john@bionbiometrics.com)



DRDC | RDDC

Canada 

# Introduction to Demographic Performance Variation

- Historically, facial recognition is known to have problems
- Google's problem with misclassified gorillas (2015)
- Congress attacks FBI for 20% error rate on NGI facial recognition. Doesn't work for coloured people (2017)
- "Gender Shades" – Problems with face detection and gender classification from IBM, Microsoft and Face++. Works for white males but not black females (2017)
- Hearings at House Oversight and Government Reform's Subcommittee on Information Technology on bias in AI systems (2018)
- Lots of press, lots of "noise"
- Is it a real issue that we need to worry about operationally?

# ISO Technical Report – Trying to Explore the Problem

- ISO/IEC Technical Report 22116 “Identifying and mitigating the differential impact of demographic factors in biometric systems”
- Editor – Canada (DRDC), Co-Editors – US and Australia
- Participating Countries
  - Australia, Canada, Finland, Germany, Italy, Israel, Japan, New Zealand, Spain, South Africa, UK, US
- Focus Areas
  - Impact of sex, age and ethnicity on iris, face, and fingerprint
- History
  - Ballot to initiate in Fall, 2016
  - First base document in January, 2017
  - Expected publication in 2020

# ISO Technical Report – Trying to Explore the Problem

- Technical Reports are not binding like standards but give advice and summarize research on newer, developing topics
- In this case, the TR is being used to raise awareness of the problem and encourage new research from participating countries
- Some agencies are undertaking new research (US NIST, Canada DRDC) and some have contributed existing research (Canada CBSA)
- Lots of information for facial recognition but much less for other modalities
- Some results seem to match with common sense
  - Fingerprint recognition is harder for older people (dry skin, arthritis, etc.)
  - Facial recognition has poor accuracy for young children (faces change a lot as you age from a baby to a toddler to a child)
- Some results are less obvious

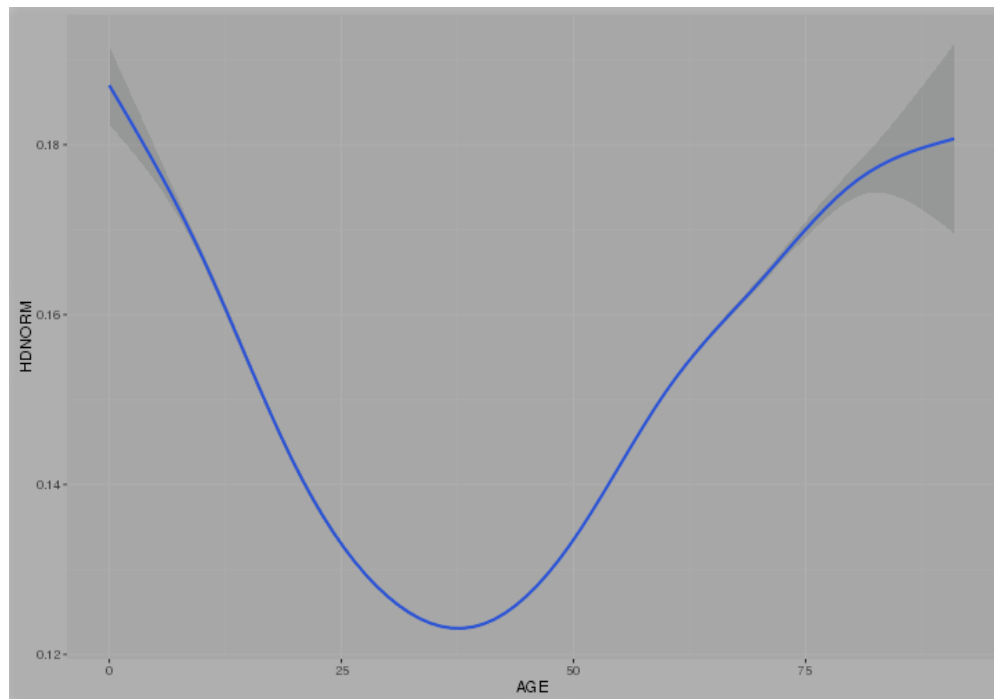
# ISO Technical Report – Summary of Results

## ■ Fingerprint Recognition

- Average quality scores for females tend to be lower than for males
- Females get rejected more than males for poor quality
- Matching accuracy for females may not be, on average, worse than for males

## ■ Iris Recognition

- Failure to Enrol rates are higher for children under 12 and adults over 70
- Hamming distances are higher (i.e. match scores are lower) for children under 12 and adults over 70



# ISO Technical Report – Summary of Results

- Facial recognition has much higher error rates for young children than for adults.
  - At a fixed threshold, false match rate (FMR) and false non-match rate (FNMR) both increase as age decreases below 18 and increase substantially for children under 4
  - FNMR seems to decrease as age increases up to about age 60 and then seems to start increasing again
- Most facial recognition algorithms perform worse for females than for males
  - At a fixed threshold many algorithms show higher FMR and FNMR for females
  - A few algorithms show higher FMR but lower FNMR or vice versa
  - No algorithms show lower FMR and FNMR for females
- But ..... What do we mean by female

# ISO Technical Report - Definitions

- Reviewing existing studies, males versus females may be called sex or gender but researchers seldom explain how this was defined
- Passports use “Sex”, ICAO says “M”, “F” or “X”
- But several countries, like Canada, New Zealand and others allow people to select their designation
- A lot of talk about gender identity but none about sex identity
- Do we mean sex or gender when we record this field
- Genetics traits and their phenotypic expressions are difficult to change
- External appearance and behaviour are easy to change
- Biometric performance is affected by both

# ISO Technical Report - Definitions

## Sex

- A classification as male, female or some other category based on an assessment of primary sexual characteristics
- Note: generally assigned at birth by a third party assessment and can only be changed with significant effort

## Gender

- A classification as male, female or some other category based on social, cultural or behavioural factors
- Note: determined through self declaration or self presentation and may change over time
- Note: depending on jurisdiction recognition may or may not require assessment by a third party



# ISO Technical Report – Summary of Results

- Facial recognition performance varies significantly with ethnicity
- At a fixed threshold, this can affect both FNMR and FMR
- FNMR can be higher by a factor of two to four
- When matched against individuals from the same ethnic group, FMR can be higher by a factor of two to twenty
- For most algorithms, Asians (including Korea, Singapore, China and Japan) have the highest FMR
- Demographic factors compound with each other
  - Sex, age and ethnicity combine to increase error rates
- When used with Automated Border Control (ABC) systems, these problems compound with ePassport quality variations

# Demographic Performance Impact – How bad can it get?

- Consider an airport with ABC gates – All numbers are examples and don't represent any real airport (but they are in a realistic range)
- Operational threshold established based on real testing with travellers in the airport
- Measured overall performance of FMR = 0.1%, FNMR = 2%
- For some nations, FNMR = 6%
  - Can be a problem for secondary manual process when large flights arrive from that nation, so staffing levels need to account for this
- Assume passport is swapped with someone of same sex, same country of nationality and similar age (plus or minus five years)
- We can evaluate this using filtered offline matching
- Now overall FMR = 5% for some nationalities
- For young children, FMR = 40%
- This represents a severe security risk

# Demographic Performance Impact – Is there a solution?

- Simplest Solution – Move to nominal FMR = 0.001%
  - But now FMR = 8% overall and for some countries is over 25%
  - Staffing levels and queue times increase significantly
- Better solution – Measure performance for each country individually with imposter matching by age, sex and nationality
  - Now the match threshold can be adjusted for each traveller based on demographic data read from the passport
  - No country has FMR > 0.1%
  - Most countries have FNMR below 2%
  - A few countries have higher FNMR, but this can be managed based on traffic flow from those countries
- Excellent solution .... IF sufficient reliable performance data available

# New Zealand eGates – Detailed Operational Performance

- NZ Customs eGates in operation since 2009
- Arrivals and departures at Auckland, Wellington, Christchurch, Queenstown
- Gates record a “video template” for each traveller
- Passport chip image and demographic details also recorded
- Data retained for a short time for performance measurement then deleted
- Offline matching system plays back video template against passport images from same traveller and a selection of other travellers to compute DET curves
- Filtering system allows imposters to be selected based on age, sex and nationality
- Regular scheduled reports of performance across all airports, all demographic groups

# New Zealand eGates – Detailed Operational Performance

- Appropriate match thresholds can then be calculated based on the NZ Customs security policy for each demographic group
- Each gate match / no match decision is based on a threshold which is adjusted based on demographic data read from passport chip
- In real operations, we have measured FMR for some groups that is almost 200 times higher than the nominal FMR
- Minimum age to use eGates is limited to 12
  - Could be as low as 10 based on performance data
  - Lower than that we start to have unacceptably high FNMR or FMR
- All the major problems associated with demographic factors have been mitigated
- This only works because detailed performance metrics were available

# Conclusions

- Facial recognition has major security issues for certain demographic groups
- They can be mitigated if sufficient detailed performance information is available and match thresholds are adjusted dynamically
- Even with dynamic thresholds, young children remain a problem
- Recommendations for ABC system deployment should now include mandatory performance testing that focuses on demographic factors and their impact on performance
- Better general guidance on measuring demographic factors in all biometric performance tests would be helpful
- ISO/IEC TR 22116 is a good reference to get started but we need more research

# DRDC | RDDC

**SCIENCE, TECHNOLOGY AND KNOWLEDGE**  
FOR CANADA'S DEFENCE AND SECURITY

**SCIENCE, TECHNOLOGIE ET SAVOIR**  
POUR LA DÉFENSE ET LA SÉCURITÉ DU CANADA

